

# STR Sequence Nomenclature: progress report from the



Green Mountain DNA Conference August 1, 2019

**STRAND** working group  
align | name | define

Katherine Gettings  
Research Biologist  
NIST Applied Genetics

**Disclaimer:**  
Identification of commercial assays and/or software is not intended to imply recommendation or endorsement by the U.S. National Institute of Standards and Technology.

# STRAND working group

align | name | define

Our mission is to harmonize related efforts across member laboratories:

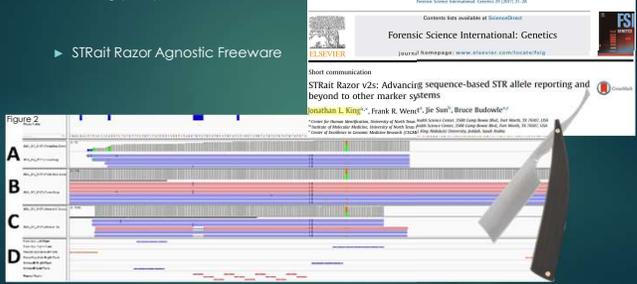
-  STRidER STR sequence quality control
-  STRSeq catalog of sequences
-  STRaitRazor bioinformatic freeware
-  Forensic STR Sequence Guide

and to characterize additional STR loci present in the genome which may be useful for forensic purposes in the future.

# STRAND working group

align | name | define

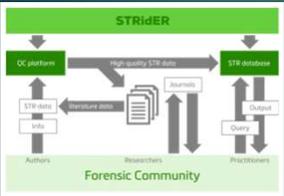
► STRait Razor Agnostic Freeware



Short communication  
STRait Razor v2s: Advancing sequence-based STR allele reporting and beyond to other marker systems  
Jonathan L. King<sup>1</sup>, Frank R. Wend<sup>2</sup>, Jie Sun<sup>3</sup>, Bruce Budowle<sup>4</sup>

# STRAND working group

align | name | define



The existing architecture of STRidER allows for the implementation of nucleotide sequence strings and thus is fully compatible with the QC of population data generated by MPS.

strider.online

# STRAND working group

align | name | define

## QC of STR sequence data on STRidER

- submission of FASTA-like strings per locus per individual
- alignment and comparison to references:  
**Forensic STR Sequence Structure Guide**  
**STRSeq catalogue**
- translation into CE allele
- inspection of flanking region where available
- **no STR sequence nomenclature** required, assessed or returned in reports
- database will host STR sequence data in future

# STR Sequence Nomenclature: Progress Report from the STRAND Working Group

align | name | define

## Forensic STR Sequence Guide

Forensic Science International: Genetics

**"The devil's in the detail": Release of an expanded, enhanced and dynamically revised Forensic STR Sequence Guide**  
G. Hillier<sup>1</sup>, K. Butler Gelling<sup>2</sup>, J.J. King<sup>3</sup>, D. Butler<sup>4</sup>, M. Bolser<sup>5</sup>, L. Borsak<sup>6</sup>, W. Pasini<sup>7</sup>

1. National Institute of Standards and Technology, Gaithersburg, Maryland, USA  
2. Forensic Science Centre, Victoria Police, Victoria, Australia  
3. University of Northumbria, Newcastle, UK  
4. University of Northumbria, Newcastle, UK  
5. University of Northumbria, Newcastle, UK  
6. University of Northumbria, Newcastle, UK  
7. University of Northumbria, Newcastle, UK

align | name | define

## Umbrella Hierarchy

24 Common auSTR Loci  
11 Alternate auSTR Loci  
26 Y-STR Loci  
7 X-STR Loci

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

Research paper  
**STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci**  
Laurence Dewar<sup>1</sup>, Jonathan King<sup>2</sup>, Waldemar Pasini<sup>3</sup>, Christopher Phillips<sup>4</sup>, Peter M. Vukobrat<sup>5</sup>, Katherine Butler Gelling<sup>6</sup>, Lisa A. Borsak<sup>7</sup>, David Bellizzi<sup>8</sup>, Martin Bolser<sup>9</sup>, Bruce Budowick<sup>10</sup>

1. National Institute of Standards and Technology, Gaithersburg, Maryland, USA  
2. Forensic Science Centre, Victoria Police, Victoria, Australia  
3. University of Northumbria, Newcastle, UK  
4. University of Northumbria, Newcastle, UK  
5. University of Northumbria, Newcastle, UK  
6. University of Northumbria, Newcastle, UK  
7. University of Northumbria, Newcastle, UK  
8. University of Northumbria, Newcastle, UK  
9. University of Northumbria, Newcastle, UK  
10. University of Northumbria, Newcastle, UK

align | name | define

## Current Sample Sets

NIST National Institute of Standards and Technology U.S. Department of Commerce

KING'S COLLEGE LONDON

UNT HEALTH SCIENCE CENTER

USC UNIVERSITY OF SOUTHERN CALIFORNIA

1786 Samples ForenSeq Verogen, PowerSeq 46GY (prototype) Promega, and GlobalFiler NGS Thermo Fisher CE Information

1043 Samples ForenSeq Verogen, CE Information

839 Samples ForenSeq Verogen, CE Information

944 Samples ForenSeq Verogen, CE Information

Working towards including more sets of data  
Collaborating with STRiER

align | name | define

## BioProject Structure

Summary: 20 per STR, 1 source accession

Format: Summary, GenBank, FASTA, SRA, RefSeq, Repeat-History, Assembly-Map, G-Map

Complete Record, Coding Sequences, Gene Features, Choose Destination, File, Collections, Clipboard

NIH STR Sequencing Project (summary)

The purpose of STRSeq is to facilitate the description of sequence-based alleles at the Short Tandem Repeat (STR) loci required for human identification assays. This collaborative effort of the international Forensic DNA Community, which has been endorsed by the executive board of the International Society of Forensic Genetics, provides a framework for communication among laboratories. Each record contains: (i) canonical sequence of an STR region; (ii) annotation of the repeat region ("breakdown") and flanking region polymorphisms; (iii) information regarding the sequencing assay and data quality; and (iv) laboratory-specific nomenclature and designations. Data within the database project is organized into locus-accessions and can be accessed by browsing, BLAST searching, or by download of files. For comments or questions, please contact [streq@nist.gov](mailto:streq@nist.gov).

Accession: PRJNA381027  
Type: Umbrella project  
Publications: 10  
Submission: 2018-04-20  
Sponsoring Organization: National Institute of Standards and Technology  
Related Resources: STRSeq, STRSeqR  
Reference: Human Identification

Resource Name	Number of Loci
Umbrella project	4

The STR Sequencing Project (summary) encompasses the following 4 sub-projects:

Project Name	Number of Projects
Umbrella project	4

Links to BioProjects

align | name | define

## Record Structure

Homo sapiens microsatellite D1S1656 14 CCTA [TCTA]13 rs1019813099 sequence

GenBank: MH174843.1

FASTA sequence: MH174843.1 | 1..138 (138bp)

STR region: 138 bp

STR Feature Features: Repeat region, Repeat unit, Repeat count

align | name | define

## Record Structure

Homo sapiens microsatellite D1S1656 14 CCTA [TCTA]13 rs1019813099 sequence

GenBank: MH174843.1

FASTA sequence: MH174843.1 | 1..138 (138bp)

STR region: 138 bp

STR Feature Features: Repeat region, Repeat unit, Repeat count







# STR Sequence Nomenclature: progress report from the

Green Mountain DNA Conference August 1, 2019

**STRAND** working group  
align | name | define

Katherine Gettings  
Research Biologist  
NIST Applied Genetics

**Disclaimer:**  
Identification of commercial assays and/or software is not intended to imply recommendation or endorsement by the U.S. National Institute of Standards and Technology.

# STRAND working group

align | name | define

Our mission is to harmonize related efforts across member laboratories:

- STRidER STR sequence quality control
- STRSeq catalog of sequences
- STRaitRazor bioinformatic freeware
- Forensic STR Sequence Guide

and to characterize additional STR loci present in the genome which may be useful for forensic purposes in the future.

# STRAND working group

align | name | define

► STRait Razor Agnostic Freeware

Short communication  
STRait Razor v2s: Advancing sequence-based STR allele reporting and beyond to other marker systems  
Jonathan L. King<sup>1</sup>, Frank R. Wendt<sup>2</sup>, Jie Sun<sup>3</sup>, Bruce Budowle<sup>4</sup>

# STRAND working group

align | name | define

Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER):  
Martin Rohlfing<sup>1</sup>, Sigi Krawinkel<sup>2</sup>, Jean-Michel Saurer<sup>3</sup>, Karl Püschel<sup>4</sup>, Peter Gill<sup>5</sup>, Dennis Goodwin<sup>6</sup>, Sören Möring<sup>7</sup>, Christophe Phillips<sup>8</sup>, Michaela Pöhl<sup>9</sup>, Peter M. Schneider<sup>10</sup>, Michaela Parson<sup>11</sup>

The existing architecture of STRidER allows for the implementation of nucleotide sequence strings and thus is fully compatible with the QC of population data generated by MPS.

strider.online

# STRAND working group

align | name | define

## QC of STR sequence data on STRidER

- submission of FASTA-like strings per locus per individual
- alignment and comparison to references:  
**Forensic STR Sequence Structure Guide**  
**STRSeq catalogue**
- translation into CE allele
- inspection of flanking region where available
- **no STR sequence nomenclature** required, assessed or returned in reports
- database will host STR sequence data in future

# STR Sequence Nomenclature: Progress Report from the STRAND Working Group

**STRAND** working group  
align | name | define

Forensic STR Sequence Guide

Forensic Science International: Genetics  
Journal homepage: [www.elsevier.com/locate/bsfr](http://www.elsevier.com/locate/bsfr)

"The devil's in the detail". Release of an expanded, enhanced and dynamically revised Forensic STR Sequence Guide  
G. Hillier<sup>1</sup>, K. Butler-Gretting<sup>2</sup>, J.J. King<sup>3</sup>, D. Butler<sup>4</sup>, M. Bolser<sup>5</sup>, L. Borsak<sup>6</sup>, W. Pasini<sup>7</sup>

1: National Institute of Standards and Technology, Gaithersburg, Maryland, USA  
2: Forensic Science International, London, UK  
3: University of Northumbria, Newcastle, UK  
4: University of Northumbria, Newcastle, UK  
5: University of Northumbria, Newcastle, UK  
6: University of Northumbria, Newcastle, UK  
7: University of Northumbria, Newcastle, UK

**STRAND** working group  
align | name | define

STRSeq BioProject

Umbrella Hierarchy

24 Common auSTR Loci  
11 Alternate auSTR Loci  
26 Y-STR Loci  
7 X-STR Loci

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

Research paper  
STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci  
Katherine Butler-Gretting<sup>1</sup>, Lisa A. Borsak<sup>2</sup>, David Bellizzi<sup>3</sup>, Martin Bolser<sup>4</sup>, Bruce Budowick<sup>5</sup>, Laurence Deveson<sup>6</sup>, Jonathan King<sup>7</sup>, Waldemar Pasini<sup>8</sup>, Christopher Phillips<sup>9</sup>, Peter M. Vallone<sup>10</sup>

**STRAND** working group  
align | name | define

STRSeq BioProject

Current Sample Sets

NIST National Institute of Standards and Technology  
KING'S COLLEGE LONDON  
UNT HEALTH SCIENCE CENTER  
USC UNIVERSITY OF SOUTHERN CALIFORNIA

1786 Samples ForenSeq Verogen, PowerSeq 46GY (prototype) Promega, and GlobalFiler NGS Thermo Fisher CE Information

1043 Samples ForenSeq Verogen, CE Information

839 Samples ForenSeq Verogen, CE Information

944 Samples ForenSeq Verogen, CE Information

Working towards including more sets of data  
Collaborating with STRiER

**STRAND** working group  
align | name | define

STRSeq BioProject

BioProject Structure

Summary: 20 per STR, 1 Source Accession

Formal: Summary, GenBank, FASTA, FASTQ, SRA, SRA History, SRA Link, SRA File, SRA Collection

Complete Record: Coding Sequences, Gene Features, Choose Destination: File, Collections, Clipboard

NIH STR Sequencing Project (summary)

The purpose of STRSeq is to facilitate the description of sequence-based alleles at the Short Tandem Repeat (STR) loci required for human identification assays. This collaborative effort of the international Forensic DNA Community, which has been endorsed by the executive board of the International Society of Forensic Genetics, provides a framework for communication among laboratories. Each record contains: (i) canonical sequence of an STR region; (ii) annotation of the repeat region ("breakdown") and flanking region polymorphisms; (iii) information regarding the sequencing assay and data quality; and (iv) laboratory-specific nomenclature and designations. Data within the database project is organized into locus sub-projects and can be accessed by browsing, BLAST searching, or by download of files. For comments or questions, please contact [streq@nih.gov](mailto:streq@nih.gov).

Accession: PRJNA381027  
Type: Umbrella project  
Publications: 10  
Submission: 2015-04-20  
Sponsoring Organization: National Institute of Standards and Technology

Resource Name	Number of Loci
Umbrella project	4

The STR Sequencing Project (summary) encompasses the following 4 sub-projects:

Project Name	Number of Projects
Umbrella project	4

**STRAND** working group  
align | name | define

STRSeq BioProject

Record Structure

Homo sapiens microsatellite D1S1656 14 CCTA [TCTA]13 rs1019813099 sequence

GenBank: MH174843.1

FASTA: MH174843.1

FASTQ: MH174843.1

STRSeq BioProject

STRSeq BioProject

**STRAND** working group  
align | name | define

STRSeq BioProject

Record Structure

Homo sapiens microsatellite D1S1656 14 CCTA [TCTA]13 rs1019813099 sequence

GenBank: MH174843.1

FASTA: MH174843.1

FASTQ: MH174843.1

STRSeq BioProject

STRSeq BioProject





